Physical Communication ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Physical Communication

journal homepage: www.elsevier.com/locate/phycom

Full length article

# Learning to short-time Fourier transform in spectrum sensing

Longmei Zhou *, Zhuo Sun, Wenbo Wang

*Beijing University of Posts and Telecommunications, Beijing, China*

A R T I C L E I N F O

*Article history:*
Received 20 April 2017
Received in revised form 10 August 2017
Accepted 12 August 2017
Available online xxxx

*Keywords:*
Deep learning
STFT
Spectrum sensing
Modulation recognition

A B S T R A C T

The future wireless communication will come up with a strict requirement on high spectral efficiency, developing novel algorithms for spectrum sensing with deep sensing capability will be more challenging. However, traditional expert feature-based spectrum sensing algorithms are lack of sufficient capability of self-learning and adaptability to unknown environments and complex cognitive tasks. To address this problem, we propose to build up a deep learning network to learn short time-frequency transformation (STFT), a basic entity of traditional spectrum sensing algorithms. Spectrum sensing based on the learning to STFT network is supposed to automatically extract features for communication signals and makes decisions for complex cognitive tasks meanwhile. The feasibility and performances of the designed learning network are verified by classifying signal modulation types in deep spectrum sensing applications.



## 1. Introduction

As the attention is now turning towards the 5th generation (5G) and beyond technologies, researchers and practitioners have joined the definition and development of the next-generation wireless networks. Though future network requires higher end-user data rates over the Internet, new traffic types and data services, energy-efficient networks are the key driving forces of the development and evolution of 5G networks. Cognitive radio (CR)which can facilitates efficient spectrum use of current licensed spectrum, is considered as a potential solution to the problem of spectrum scarcity. However, there are further challenges for spectrum sensing technologies in pace with developing of future 5G networks. Firstly, spectrum sensing should be developed toward deep sensing, which means that in addition to detect spectrum occupancy, it is also necessary to exploit and acquire the comprehensive characteristics of spectrum and the occupied signals to meet strict requirements of spectrum efficiency. On the other hand, the capability of self-learning and self-decision making under unknown environmental is necessary for the future communication systems or equipments to handle varied signal types, complex networks and cognitive tasks. Due to the innate constraints of framework based on expert knowledge and predefined mathematical models, traditional spectrum sensing methods cannot essentially confront and solve these challenges. New methods of spectrum sensing should be developed. One of the potential breakthrough point is to make systems or equipments capable of learning to spectrum sensing from data and signal. In this paper, we will try to build up a neural network with the core of learning to STFT in maintaining time-frequency representations in spectrum sensing applications.

As a matter of fact, many scholars have tried to incorporate machine learning into cognitive communications. It is pointed out that machine learning can solve some problems in large scale MIMO, D2D network, heterogeneous network, cognitive radio and so on [1]. The current results of related researches can be divided into two categories. One applies neural network or shallow learning network (such as SVM, K-means) based classifiers on expert features for recognition in different scenarios [2–6]. The other one uses some mature deep learning models in natural language processing or image processing to the cognitive communication fields for feature extraction. There are quite few attempts to remodeling learning networks based on the communication characteristics, but basically applying shallow network based models to the problems encountered in communications.

Inspired by the works of Timothy J. O'Shea and other researchers [7], in which they proposed a theory of learning to communicate which re-modeled the transmission, reception and synchronization of communication with framework of neural network, this paper tries to build up a deep learning network that learns to STFT in spectrum sensing. The learning network is actually a linear fitting of STFT with particular neural layers and connections. The learning to STFT network, as a novel concept and framework of spectrum sensing, is proved capable to cope with various cognitive scenarios with complex signal types and complex networks. The feasibility of the proposed learning network for deep spectrum sensing is verified by through the application on modulation recognition tasks.

The rest of this paper is organized as follow. As the preliminary of main contributions, Section 2 briefly reviews the time-frequency

* Corresponding author.
*E-mail address:* 173291791@qq.com (L. Zhou).

http://dx.doi.org/10.1016/j.phycom.2017.08.007
1874-4907/

analysis methods and the short-time Fourier transform at first, then typical layers in neural networks are reviewed. Section 3 proposes the network architecture of the learning to STFT network and illustrates each operation layers respectively. The approach and related issues for network training are also analyzed in this section. In Section 4, as a test of proposed method in spectrum sensing application, the problem of recognizing three basic types of digital modulation signals are set up and simulated. The overall performance of the network is analyzed from the aspects of classification accuracy.

## 2. Problem formulation

### 2.1. Limitation of traditional time-frequency analysis

Joint time-frequency transforms were developed for the purpose of characterizing time-varying frequency contend of a signal since one-dimensional solution is not sufficient in some cases. Time-frequency analysis is a form of local Fourier analysis that treats time and frequency simultaneously and symmetrically. The well-known time-frequency representation of time signal dates back to Gabor in 1993 [8], which is known as STFT and has been extensively applied to signal analysis, communication theory, and image processing, etc.

The basic idea, or the most standard approach of STFT is to decompose the time-domain signal into many segments with a window and then calculate the Fourier transform (or the frequency spectrum) of each segment. Consider a discrete signal $\mathbf{x} \in R^N$ and a discrete sampled window function $\gamma \in R^M$. To obtain a localized spectrum of $\mathbf{x}$ at time $n$, the signal is decomposed with the window $\gamma$ centered at time $m = n$ and take the Fourier transform w.r.t $m$, the STFT transform $F(n, \omega)$ then can be defined as

$$F(n, \omega) = \sum_{m} x(m)\gamma(m-n)e^{-jn\omega} \tag{1}$$

The magnitude squared of the STFT is called the spectrogram, which provides time-localized frequency content of the transformed signal. The spectrogram can be written as

$$S(n, \omega) = |F(n, \omega)|^2 \tag{2}$$

The fixed window size of STFT limits its ability to span both time and frequency of unknown signals with resolution well matched the signal characteristics [9]. A large window results in better frequency resolution, but leads to worse time resolution, and vice versa. When the frequency content of a signal or channel characteristics change rapidly with time, a small time window is necessary, but frequency resolution is degraded, while a large window size may be sufficient when the frequency characteristics change slowly. As for unknown characteristics, a few attempts of different window sizes will be required to find ideal time and frequency resolution.

Multi-resolution analysis (MRA), e.g. Wavelet transform, is an improved method in case of the resolution limitation of STFT. MRA analyzes the signal at different frequencies with different resolutions by employing orthonormal filters consisting of low pass filter and multiple band pass filters [10,11], and gives good time resolution and poor frequency resolution at high frequencies, or good frequency resolution and poor time resolution at low frequencies. And many different conventional filters can be used to implement it.

### 2.2. Feasibility of learning to STFT

This paper seek to build up a deep learning network to learn STFT, which can automatically extract features of communication signals, by referring to the processing procedures of STFT. Based on this objective, we firstly decomposes the procedure of STFT into three steps in brief: convolutes the signal with window function step by step, sub-samples the results with a stride of half of window size, then normalizes the results. What follows is a brief review of the similar operations in neural networks.

- **Convolution**: This is a basic operation in neural networks such as CNN, convolution auto-encoder (CAE), convolutional restricted Boltzmann machine (CRBM), etc. The visible values ($v$) are convoluted by a shared weights ($W$) and biased with a shared bias value $b$ within a group that the raw values of hidden units are $W * v + b$. The number of hidden units depends on the length of $v$ and $W$ and the convolution mechanism (narrow or wide).
- **Pooling**: The three most commonly used pooling methods are general pooling (e.g. mean pooling, max pooling), overlapping pooling [12] and spatial pyramid pooling [13]. Regardless of the specific meaning of these methods, one of their most fundamental effects is the sub-sampling of signals. Pooling layer hence is also called as sub-sampling layer sometimes.
- **Normalization**: The normalization of deep learning is reflected in two aspects. First, the normalization of the input values, where the range of input data will effect the initialization of parameters. Second, the normalization of the output values, where too big value will bring a numerical problem to the update of gradients and the setting of initial learning rate.

According to hereinbefore reviews, it is feasible to build up a neural network which learns the process of STFT with convolutional layer, sub-sampling layer and normalization layer.

## 3. Learning to STFT

An illustration of the designed network architecture is shown in Fig. 1. The original input are real-valued $2 \times N_v$ vectors, layer "Conv1" contains $K_1$ one-dimensional filters of length $N_w$. Raw convolutional results of each group, sized at $1 \times (N_v - N_w + 1)$, are feed into sub-sampling layer that elements for every $N_w/2$ steps will be kept in the output for each group. Output vectors of each group are combined as a matrix (image) sized $K1 \times N_{v2}$ ($N_{v2} = \lceil 2(N_v - N_w + 1)/N_w \rceil$). Layer "Conv2" is identical to Conv1, but contains $K_2$ filters with $K_1$ channels, in other words, the filters in Conv2 are matrices sized at $K_1 \times N_{w2}$. A simple summary statistics, e.g mean, is used for each channel over all the frames (that is the max-pooled Conv2-layer activations) and achieves a $1 \times K_2$ feature vector for each signals.

### 3.1. Convolutional layer

The Conv1 layer and Conv2 layer are both CRBM practically. The CRBM is an extension of RBM [14] to a convolutional setting, in which the weights between the hidden units and the visible units are shared among all locations in the hidden layer. The learning to STFT network is supposed to be separated from deeper feature extraction networks, as a result, CRBM is chosen since they are trained separately.

The major difference between Conv1 and Conv2 is the input channel size, where Conv1 has a channel of 2, which is determined by the two-channels real-valued time-series input data, while Conv2 has a channel size that equals to the number of filters in

Please cite this article in press as: L. Zhou, et al., Learning to short-time Fourier transform in spectrum sensing, Physical Communication (2017), http://dx.doi.org/10.1016/j.phycom.2017.08.007.